

Sample: ERR2383627

FASTQ Quality

| Filename | ERR2383627_R1.fastq.gz | ERR2383627_R2.fastq.gz |
|---------------------|------------------------|------------------------|
| File Size | 344.0 MB | 385.2 MB |
| Q30 Passing | 85.8% | 67.74% |
| Mean Read Score | 27.5 | 22.4 |
| Average Read Length | 230.1 | 230.1 |

Read Mapping against ERR2383627

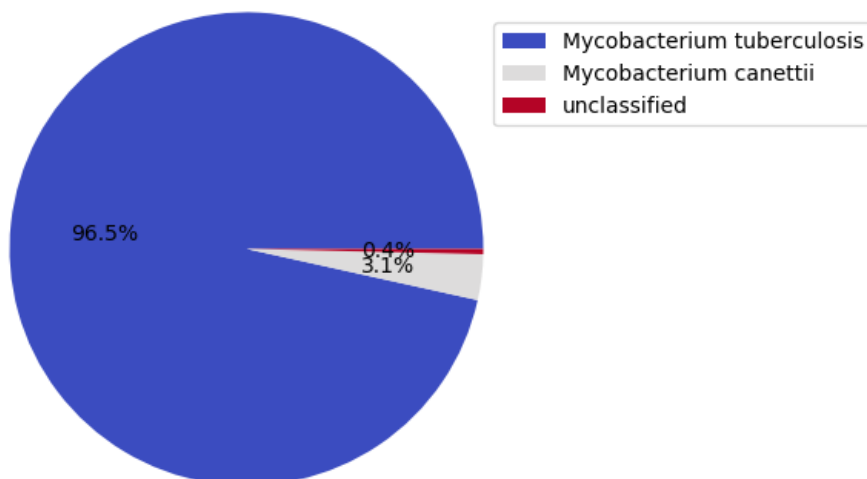
| | | | | | |
|------------------------|------------------------|-----------------------------------|------------------|----------------------------|--------------|
| Mapped Paired Reads | Mapped Single Reads | Unmapped Reads | Unmapped Percent | Unmapped Assembled Contigs | |
| 1,990,131 | 14,415 | 17,173 | 0.8% | assembly not done | |
| Duplicate Paired Reads | Duplicate Single Reads | Duplicate Percent of Mapped Reads | | | |
| 36,352 | 9,086 | 2.0% | | | |
| BAM File | Reference Length | Genome with Coverage | Average Depth | No Coverage Bases | Quality SNPs |
| ERR2383627_nodup.bam | 4,489,898 | 99.94% | 192.0X | 2,478 | 250 |

Assembly

| | | | | | |
|--------------|--|----------------|--------------|--------|--------------------------------|
| Contig count | Contig length counts < 301-999bp > | Longest contig | Total length | N50 | FASTQ calculated mean coverage |
| 307 | 18 188 101 | 204,921 | 4,489,898 | 79,426 | 205.6X |

FASTQ Identifications

FASTQ Read Identification



Identified using: Kraken and Bracken

Genome Size Comparisons

| | |
|----------------------------|-------------------------|
| Mycobacterium tuberculosis | expected: 4,075,000 nt |
| ERR2383627 | assembled: 4,489,898 nt |

Sourmash Sequence Similarity

| Similarity | ID |
|------------|--|
| 97.0% | KK338934.1 Mycobacterium africanum MAL010118 genomic scaffold adPbt-supercont1.1, whole genome shotgun sequence |
| 96.9% | KK338573.1 Mycobacterium tuberculosis CPHL_A genomic scaffold adCAM-supercont1.1, whole genome shotgun sequence |
| 96.8% | KK338837.1 Mycobacterium africanum MAL010074 genomic scaffold adOUR-supercont1.1, whole genome shotgun sequence |
| 96.3% | KK339435.1 Mycobacterium tuberculosis OFXR-22 genomic scaffold adiej-supercont1.1, whole genome shotgun sequence |
| 96.2% | KK353227.1 Mycobacterium tuberculosis TKK_03_0029 genomic scaffold adrRI-supercont1.1, whole genome shotgun sequence |
| 96.2% | KK322458.1 Mycobacterium tuberculosis M1300 genomic scaffold adOXC-supercont1.1, whole genome shotgun sequence |
| 96.2% | KK328761.1 Mycobacterium tuberculosis M1424 genomic scaffold adPcq-supercont1.1, whole genome shotgun sequence |
| 96.2% | KK321818.1 Mycobacterium tuberculosis M1026 genomic scaffold adPgg-supercont1.1, whole genome shotgun sequence |
| 96.2% | KK321735.1 Mycobacterium tuberculosis M1022 genomic scaffold adPai-supercont1.1, whole genome shotgun sequence |
| 96.1% | KK330835.1 Mycobacterium tuberculosis M1928 genomic scaffold adOYM-supercont1.1, whole genome shotgun sequence |

MLST 2.19.0

Schema-Sequence type: **mycobacteria-**

MLST Detail: S14Z(17), L35(15), S19(20), L19(159), S12(145), S8(18), L16(22), S7(19)

Results provided by: MLST

AMRFinderPlus - version 3.11.2, database 2021-03-01.1

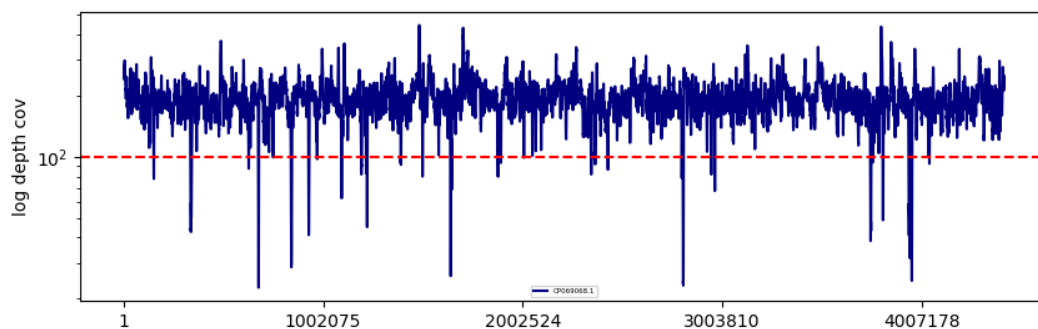
AMRFinderPlus Organism Option Not Found: Mycobacterium tuberculosis

| Contig ID | Element | Gene | Description | % Cov | % Similar |
|------------------------------------|---------|------------|---|--------|-----------|
| NODE_32_length_61418_cov_94.317273 | AMR | erm(37) | 23S rRNA (adenine(2058)-N(6))-methyltransferase Erm(37) | 100.0% | 100.0% |
| NODE_52_length_31008_cov_78.629708 | AMR | blaC | class A beta-lactamase BlaC | 100.0% | 100.0% |
| NODE_55_length_26276_cov_81.948602 | AMR | aac(2')-Ic | aminoglycoside N-acetyltransferase AAC(2')-Ic | 100.0% | 100.0% |

Results provided by: AMRFinderPlus

Coverage Graph

ERR2383627 reads aligned against CP069068



| ID | Average Depth | Genome Coverage | Genome Length | Ambiguous SNPs |
|------------|---------------|-----------------|---------------|----------------|
| CP069068.1 | 193.4X | 99.8% | 4,420,832 | 116 |

ID Detail

CP069068.1 Mycobacterium tuberculosis strain N1272 chromosome, complete genome

BLAST nt - Assembly Identification

| Nucleotide Representa- tion | Description |
|-----------------------------------|--|
| 1,061,397 | CP069068 Mycobacterium tuberculosis strain N1272 chromosome, complete genome |
| 1,048,982 | CP069069 Mycobacterium tuberculosis strain N1176 chromosome, complete genome |
| 573,015 | CP058236 Mycobacterium tuberculosis strain 060827 chromosome, complete genome |
| 470,431 | CP069067 Mycobacterium tuberculosis strain N1268 chromosome, complete genome |
| 146,028 | CP014617 Mycobacterium africanum strain UT307 genome |
| 145,860 | CP025599 Mycobacterium tuberculosis strain GG-45-11 chromosome, complete genome |
| 132,960 | CP130806 Mycobacterium tuberculosis strain MTb-Oman-3211453 chromosome, complete genome |
| 116,972 | CP096844 Mycobacterium tuberculosis variant bovis strain Mb1855 chromosome, complete genome |
| 92,139 | CP023631 Mycobacterium tuberculosis strain TBDM1506 chromosome, complete genome |
| 88,858 | CP041809 Mycobacterium tuberculosis strain SEA09020048P6C4 chromosome, complete genome |
| 86,359 | CP041857 Mycobacterium tuberculosis strain 1-0069P6C4 chromosome, complete genome |
| 84,538 | CP003494 Mycobacterium bovis BCG str. ATCC 35743, complete genome |
| 83,944 | CP028428 Mycobacterium tuberculosis strain CAS chromosome |
| 67,040 | CP046308 Mycobacterium tuberculosis strain FDAARGOS_751 chromosome, complete genome |
| 64,301 | CP046507 Mycobacterium avium subsp. avium strain DSM 44156 chromosome, complete genome |
| 55,961 | CP048071 Mycobacterium tuberculosis strain RW-TB008 chromosome, complete genome |
| 39,864 | CP069063 Mycobacterium tuberculosis strain N3913 chromosome, complete genome |
| 36,269 | LR882498 Mycobacterium tuberculosis variant microti strain Mycobacterium microti Maus III genome ass |
| 25,591 | CP072790 Mycobacterium tuberculosis strain I0003447-5 chromosome, complete genome |
| 16,094 | CP054013 Mycobacterium tuberculosis strain FDAARGOS_757 chromosome, complete genome |
| 13,254 | CP054014 Mycobacterium tuberculosis strain FDAARGOS_756 chromosome, complete genome |
| 6,761 | CP025604 Mycobacterium tuberculosis strain GG-129-11 chromosome, complete genome |
| 5,513 | LT703009 Andersenella sp. Alg231_50 genome assembly, chromosome: VII |
| 5,126 | CP033310 Mycobacterium tuberculosis variant bovis BCG strain BCG-S48 chromosome |
| 2,490 | CP053068 Mycobacterium avium subsp. paratuberculosis strain DSM 44135 chromosome, complete genome |
| 1,647 | AP022597 Mycobacterium paraintracellulare JCM 30622 DNA, complete genome |
| 1,364 | CP130792 Mycobacterium tuberculosis strain MTb-Oman-3213195 chromosome, complete genome |
| 1,128 | CP047258 Mycobacterium tuberculosis strain TCDC3 chromosome |
| 886 | CP009191 Mycobacterium tuberculosis TRS24 genome |
| 885 | CP016396 Mycobacterium avium strain RCAD0278 isolate SC150824 chromosome, complete genome |
| 877 | AP019684 Xanthomonas campestris pv. campestris MAFF302021 DNA, complete genome |
| 744 | CP046309 Mycobacterium tuberculosis strain FDAARGOS_750 chromosome, complete genome |
| 636 | CP096843 Mycobacterium tuberculosis variant bovis strain Mb3602 chromosome, complete genome |
| 605 | CP025603 Mycobacterium tuberculosis strain GG-121-10 chromosome, complete genome |
| 592 | LR699570 Mycobacterium tuberculosis variant bovis strain Mb3601 genome assembly, chromosome: Mb3601 |
| 525 | CP060407 Mycobacterium avium subsp. hominissuis strain w14 chromosome |
| 472 | CP045963 Mycobacterium intracellulare subsp. chimera strain AUSMDU00007395 chromosome, complete gen |
| 442 | CP118560 Enterobacter kobei strain 2020CK-00216 chromosome |
| 434 | CP025607 Mycobacterium tuberculosis strain GG-186-10 chromosome, complete genome |
| 429 | MN045317 Synthetic construct clone phix_A1_A2_A3_A4_deopt DNA replication initiation protein (A |
| 425 | CP130791 Mycobacterium tuberculosis strain MTb-Oman-3213249 chromosome, complete genome |
| 417 | CP005082 Mycobacterium tuberculosis str. Beijing/NITR203, complete genome |
| 410 | MN045327 Synthetic construct clone phix_F2_deopt capsid protein (F) gene, complete cds |
| 409 | CP014566 Mycobacterium tuberculosis variant bovis BCG str. Tokyo 172 chromosome, complete genome |
| 409 | CP069072 Mycobacterium tuberculosis strain N1216 chromosome, complete genome |
| 404 | LR828289 Ananas comosus genome assembly, chromosome: 9 |
| 401 | MN045340 Synthetic construct clone phix_H2_H3_deopt minor spike protein (H) gene, complete cds |
| 395 | CP011008 Bacillus simplex strain SH-B26, complete genome |
| 392 | FO203508 Mycobacterium canettii CIPT 140070008 complete genome |
| 385 | CP130782 Mycobacterium tuberculosis strain MTb-Oman-3213573 chromosome, complete genome |
| 383 | EF380015 Enterobacteria phage phiX174 isolate 10C90, complete genome |
| 249 | CP041848 Mycobacterium tuberculosis strain 1-0137P6C4 chromosome, complete genome |
| 249 | CP069075 Mycobacterium tuberculosis strain N0004 chromosome, complete genome |
| 196 | CP053092 Mycobacterium tuberculosis strain 4860 chromosome, complete genome |

Results provided by: BLAST nt database

Report Description

Input to create this report was either FASTA, or single/paired FASTQ/s.

If built from FASTA, contigs have been identified using Kraken2/Bracken; and assembly metrics, expected genome size compared against total assembly length, MLST, PlasmidFinder, AMRFinderPlus and BLAST results have been reported. In addition to these tests, if report was built from FASTQ/s, sequence technology, FASTQ metrics and coverage graph against the top BLAST hit was also reported. Finally, if Kraken2/Bracken top finding was Salmonella SeqSero2 and CRISPR SeroSeq additionally reported. See table below for summary of tests reported.

Expected results are approximately based on a 4,500,000 base bacterial genome.

When tests are run and there are no significant findings the test's banner and a "no results" message will still be included in this report. If this is not seen, yet test is expected, the report is invalid.

| | FASTA | FASTQ | FASTQ Identified as Salmonella |
|----------------------|-------|-------|--------------------------------|
| Sequence Technology | | X | X |
| FASTQ metrics | | X | X |
| Kraken2/Bracken | X | X | X |
| Assembly metrics | X | X | X |
| Expected Genome Size | X | X | X |
| MLST | X | X | X |
| PlasmidFinder | X | X | X |
| AMRFinderPlus | X | X | X |
| Coverage Graph | | X | X |
| BLAST | X | X | X |
| SeqSero2 | | | X |
| CRISPR SeroSeq | | | X |

Table 1: Test Summary

Sequence Technology

Single FASTQ is assumed to have been generated from Ion S5 System using S5 chemistry. Paired reads are assumed to have been generated from Illumina MiSeq System. Read length is reported based on average read length and rounded up to 50th number.

FASTQ metrics

Filename: Input files

File Size: Compressed input file size (.gz). *Expected: >150MB*

Q30 Passing: Percentage of reads with an average Phred score above 30. *Expected: >70%*

Mean Read Score: Average Phred score of all base calls. *Expected: >30*

Kraken2/Bracken

Bracken computes specie abundance using Kraken2 taxonomy labels. Fraction of reads are reported only when greater than 1% of total reads.

Kraken2 database was built from the following:

Libraries:

archaea: RefSeq complete archaeal genomes
 bacteria: RefSeq complete bacterial genomes
 plasmid: RefSeq plasmid nucleotide/protein sequences
 viral: RefSeq complete viral genomes
 human: GRCh38 human genome
 fungi: RefSeq complete fungal genomes
 plant: RefSeq complete plant genomes
 protozoa: RefSeq complete protozoan genomes
 UniVec Core: A subset of UniVec chosen to minimize false positive hits to the vector database

Custom genomes:

Bos taurus (cow)
 Sus scrofa (pig)
 gallus gallus (chicken)
 Odocoileus virginianus (white-tailed deer)
 Panthera tigris altaica (tiger)
 Felis catus (house cat)
 Canis lupus familiaris (dog)
 Loxodonta africana (elephant)
 Tursiops truncatus (bottlenose dolphin)
 Anas platyrhynchos (mallard)
 Cyprinus carpio (common carp)
 Nanorana parkeri (frog)
 Aedes albopictus (mosquito)

Unclassified reads were either poor quality or no perfect sample k-mers (k=35) mapped to read sequence from the Kraken2 database build.

Assembly metrics

Reads are assembled using SPAdes when FASTQs are used to generate report. When a FASTA is used SPAdes is assumed assembler.

Metrics can vary greatly. Numbers outside expected values should not always be interpreted as a poor sample, but should be considered against other test results.

Scaffolds: Total contig count of assembly. *Expected: <300*

Total Length: Total length of all sequences. *Expected: 4,500,000*

Longest Scaffold: Longest assembled sequence. *Expected: >200,000*

Scaffolds >1K nt: Sequences greater than 1,000 nucleotides in length. *Expected: >200*

N50: Half of the total assembled length is made from sequence greater than this value. *Expected: >100,000*

Mean Read Depth: Estimated average read depth across total assembly length. *Expected: >50X*

Genome Size Comparisons

Expected genome size is found using top Bracken identification. The genome size is estimated by providing the species name to NCBI which returns the expected size. If NCBI does not return a genome size for the top Bracken identification the Bracken list is iterated from highest to lowest read identification until genome size is found. If an expected genome size is still not found a value of 5,000,000 is used.

The comparison between sample assembly size and expected genome size is very important when determining contamination and genome completeness.

MLST

Assembly file scanned for PubMLST typing schemes using MLST tool.

Schema-Sequence type: Traditional PubMLST typing scheme found

MLST Detail: Gene and allele IDs, see link for explanations for "n", "~", "?", "-" and "n, m" notation.

PlasmidFinder

PlasmidFinder tool used to target plasmids.

Description: NCBI accession number and identification

Length: Sequence length of described accession

Contig ID: Assembly identification of found plasmid

HSP Length: High score pairs length of BLAST database sequence identifying plasmid

% Identity: Percentage of HSP bases similar to matched findings

% Coverage: Percentage of HSP bases aligning to matched findings

AMRFinderPlus

AMRFinderPlus finds acquired antimicrobial resistance genes in assembled nucleotide sequences and resistance-associated point mutations in select taxa. Gene classes found include AMR, point mutations, virulence factors, biocide, heat, acid, and metal resistance genes.

This report includes providing an organism option to AMRFinderPlus if Bracken identification fits select options. When an organism option is provided point mutations are screened and some common findings are suppressed.

AMRFinderPlus Organism Option: Option used to get organism-specific results

Contig ID: Assembly identification of found element

Element: Classification of the AMRFinderPlus gene

Gene: AMR gene

Name: Gene detail

% Cov: Percentage of protein aligning to matched findings

% Similar: Percentage of protein similar to matched findings

Coverage Graph

Reads aligned using BWA to the top BLAST accession hit. When greater than 100X depth-of-coverage log read depth is graphed. If less than 100X a linear read depth is graphed.

ID: NCBI accession aligned against

Average Depth: Average depth-of-coverage across sequence aligned against

Genome Length: Length of genome aligned against

Ambiguous SNPs: Ambiguous SNP count. A high count can indicate poor WGS (i.e. low coverage), a multiple-strain sample or contamination with a closely related strain.

ID Detail: Accession detail

BLAST

Assembly BLAST against nt database. Totaled nucleotide representation identified as top hit.

Nucleotide Representation: Summed contig lengths based on same accession top hit identifications.

Description: NCBI accession number and identification

SeqSero2

SeqSero2 targets the O (wzx and wzy genes) and H (fliC and fljB genes) antigens to determine serotype. When more than one O antigen, or two or more H antigens are found it can indicate the presence of contamination.

Predicted serotype(s): Predicted serotype based on O and H antigens

Predicted antigenic profiles: Shorthand profiles

Predicted subspecies: Predicted subspecies if available

Note: Special notes from SeqSero when provided

Bottom: Contig detail of SeqSero findings